

ABSTRACT

5

BASECALLING SYSTEM AND PROTOCOL

Using data electrophoretic trace data from conventional nucleic acid sequencing equipment, a method for basecalling that is tolerant to variable peak spacing is described. The method generates high-quality basecalls and reliable quality scores. In addition, a new type of quality score that estimates the probability of a deletion error between the current and the following basecall is described. A new protocol for benchmarking that better discerns basecaller performance is also provided.